

בנוס – בניית מודל ML חדש:

הקדמה:

בחרנו להשתמש במודל RandomForest לשם החיזוי. בחרנו להשתמש ב10 עצים במודל (numtrees=10).

המודל ידוע לחיוב בהתמודדות שלו עם שונות בנתונים, מה שמשקף בצורה טובה את המצב בנתונים שלנו. חשוב לציין כי המודלים נבדלים בשלב preprocess בלבד.

Model flow using Pipeline

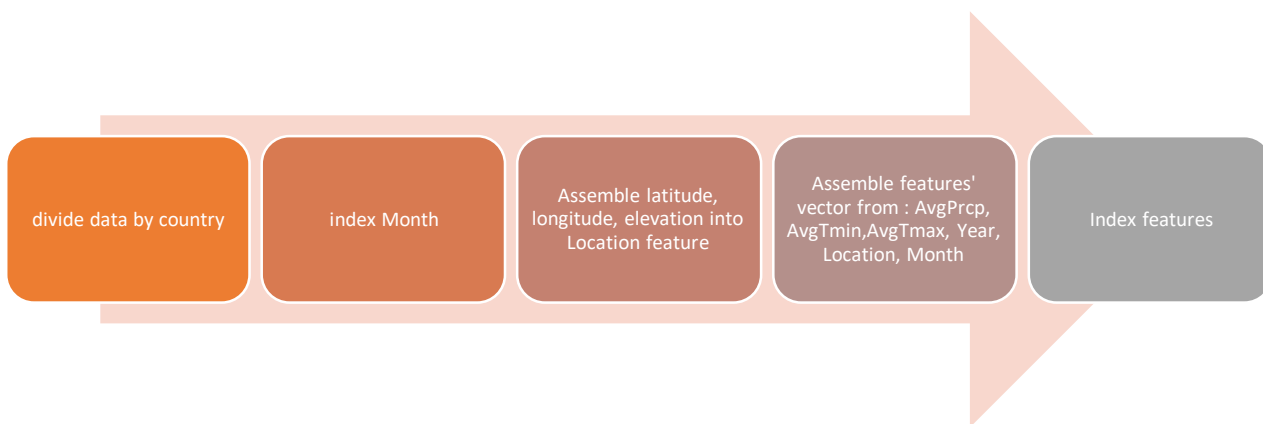


בהתאם לכך נקטנו בגישות הבאות:

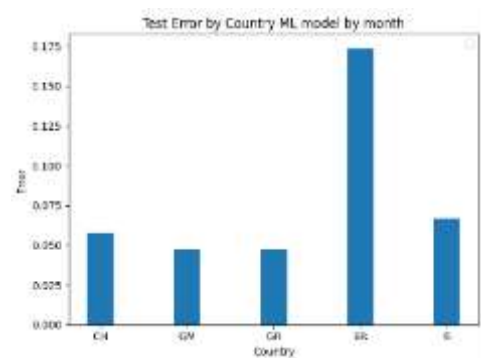
גישה 1: חיזוי חודשים

ניסינו לבנות מודל שחזוה את החודש שמתאים לדאטא לכל מדינה. בשל השונות הגבוהה של הנתונים בחודשים, קיבלנו אחוזי דיוק נמוכים ולכן ניסינו לשפר את המודל בגישה 2.

Pre process Data:



תוצאות:



נשים לב כי השגיאה פחות או יותר זהה בין המדינות חוץ מאשר ברזיל. ברזיל, מדינה בה מזג אוויר חריג ושונה משאר המדינות. לכן נשער כי המודל שלנו מתאים למדינות בהן מזג האוויר לא טרופי. מבין ארבעת המדינות שהן לא ברזיל, קל לראות כי השגיאה עבור ישראל היא הגבוהה ביותר. אנחנו משערים כי בישראל אקלים דומה לאורך כל השנה, כאשר האקלים הוא חם וצחיח ולאורך השנים, ידועות תקופות חריגות של בצורות ומזג אוויר לא אופייני, אשר כנראה השפיע בשלב האימון.

גישה 2: חיזוי עונות

בחרנו לנסות לחזות את העונה המתאימה לדאטא בחלוקה למדינות זאת על מנת ליצור מודל כללי יותר. מאחר והשונויות של הנתונים גבוהה מאד בין מדינות, בחרנו לבצע את החיזוי לכל מדינה בנפרד. כמו כן, מאחר והמדינות שלקחנו נמצאות ביבשות שונות, חילצנו מהאינטרנט את החודשים המתאימים לכל עונה ויצרנו מילון שמכיל את החודשים שמתאימים לכל עונה בכל מדינה.

הקשר בין חודש-עונה-מדינה:

סימוכין להתאמה בין עונה, מדינה וחודשים:

סין ואירופה:

SEASONS OF THE YEAR AND NORTH	SEASONS AND MONTHS AROUND THE WORLD							
	SUMMER	DRY SEASON	RAINY SEASON	AUTUMN	WINTER	DRY SEASON	RAINY SEASON	SPRING
Europe	June July August			September October November	December January February			March April May
North America	June July August			September October November	December January February			March April May
India	April May June		April May June July August	July August September	October November December		November December January February March	February March
China	May June July			July August September October	November December January February			March April

עונות בישראל : <https://seasonsyear.com/Israel#content-1>

עונות בברזיל: <https://rioandlearn.com/seasons-in-brazil>

Heat map by season:

SUMMER **SPRING**

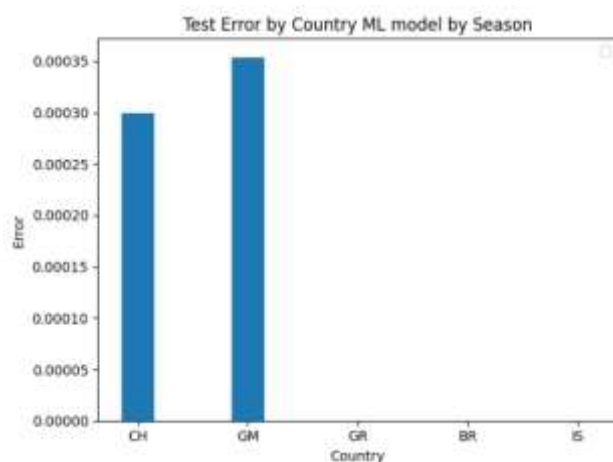
AUTUMN **WINTER**

brazil	israel	greece	germany	china	country/Month
					1
					2
					3
					4
					5
					6
					7
					8
					9
					10
					11
					12

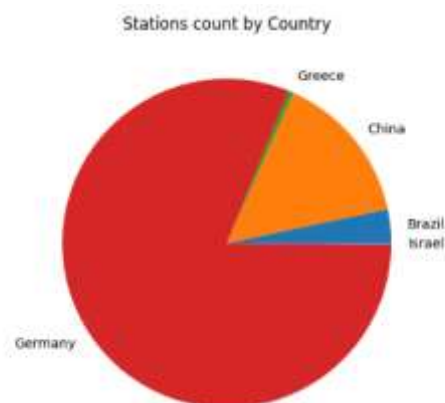
Pre process Data:



תוצאות:



נשים לב כי השגיאה במדינות יוון, ברזיל וישראל היא אפס. כמו כן, השגיאה בגרמניה היא הגבוהה ביותר מבין סין וגרמניה. נזכר בהתפלגות כמות התחנות שמשיכות לכל מדינה :



נשים לב, באופן מפתיע השגיאה של סין וגרמניה היא הגבוהה ביותר וכן כי כמות התחנות שמשיכות אליהן היא הגבוהה ביותר. השערה לכך היא שמזג האוויר במדינות סין וגרמניה חריג יותר לאורך השנה וכי העונות לא מופרדות בצורה ברורה (לדוגמה : גשם לאורך כל השנה שלא מאפיין עונה ספציפית). זאת ניתן גם לקשר לאנליזה שביצענו "קשר בין מיקום וזמן" – שם ראינו כי בסין יש קפיצות חדות לאורך השנה. נוכל לשער כי קפיצות אלו משתנות לאורך השנים ולכן קשות לחיזוי.

לעומת זאת, יוון וישראל הן מדינות דומות לכן באופן לא מפתיע השגיאה של חיזוי העונות זהה ואפסית.

